Available online at www.sciencedirect.com

SCIENCE DIRECT®

International Journal of Approximate Reasoning
42 (2006) 21–36

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

www.elsevier.com/locate/ijar

Operations for inference in continuous Bayesian networks with linear deterministic variables

Barry R. Cobb [a,*], Prakash P. Shenoy [b]

[a] *Department of Economics and Business, Virginia Military Institute, Lexington, VA 24450, USA*
[b] *School of Business, University of Kansas, 1300 Sunnyside Ave., Summerfield Hall, Lawrence, KS 66045-7585, USA*

Available online 15 November 2005

Abstract

An important class of continuous Bayesian networks are those that have linear conditionally deterministic variables (a variable that is a linear deterministic function of its parents). In this case, the joint density function for the variables in the network does not exist. Conditional linear Gaussian (CLG) distributions can handle such cases when all variables are normally distributed. In this paper, we develop operations required for performing inference with linear conditionally deterministic variables in continuous Bayesian networks using relationships derived from joint cumulative distribution functions. These methods allow inference in networks with linear deterministic variables and non-Gaussian distributions.

Keywords: Bayesian networks; Conditional linear Gaussian models; Deterministic variables

1. Introduction

Bayesian networks model knowledge about propositions in uncertain domains using graphical and numerical representations. At the qualitative level, a Bayesian network is a directed acyclic graph where nodes represent variables and the (missing) edges represent conditional independence relations among the variables. At the numerical level, a Bayesian

* Corresponding author. Fax: +1 540 464 7005.
E-mail addresses: cobbbr@vmi.edu (B.R. Cobb), pshenoy@ku.edu (P.P. Shenoy).

0888-613X/$ - see front matter
doi:10.1016/j.ijar.2005.10.002

network consists of a factorization of a joint probability distribution into a set of conditional distributions, one for each variable in the network. Continuous Bayesian networks contain variables whose state spaces are uncountable.

A commonly used type of Bayesian network which accommodates continuous variables is the conditional linear Gaussian (CLG) model [5,7]. In CLG models, the distribution of a continuous variable is a linear Gaussian function of its continuous parents. The scheme originally developed by Lauritzen [7] allowed exact computation of means and variances in CLG networks when the conditional distribution of a variable given its continuous parents has a positive variance; however, this algorithm did not always compute the exact marginal densities of continuous variables. A new computational scheme for CLG models was developed by Lauritzen and Jensen [8]. To find full local marginals, this scheme places some restrictions on the construction and initialization of junction trees.

An important class of continuous Bayesian networks are those that have linear conditionally deterministic variables (a variable that is a deterministic function of its parents). In this case, the joint density function for the variables in the network does not exist. CLG models can handle such cases when all variables are normally distributed. However, for models where continuous variables are not normally distributed, methods for carrying out exact inference in networks with linear deterministic relationships have not been developed.

Exact inference in hybrid Bayesian networks can be performed using mixtures of truncated exponential (MTE) potentials [9,12]. General formulations of MTE potentials which approximate the normal probability density function (PDF) exist [1]; however, these formulations cannot be used to model a conditional distribution where the variance of a variable given values of its continuous parents is zero. In this paper, we develop inference operations for linear conditionally deterministic variables using relationships derived from joint cumulative distribution functions (CDFs). These operations allow MTE potentials to be used for inference in any continuous CLG model, as well as other models that have linear conditionally deterministic variables which are non-Gaussian.

The rest of this paper is organized as follows. Section 2 introduces notation and definitions used throughout the paper. Section 3 introduces techniques for using CDFs to construct PDFs for deterministic variables. Section 4 introduces join tree operations for linear deterministic variables. Section 5 contains an example of inference in a continuous Bayesian network containing linear deterministic variables. Section 6 summarizes and states directions for future research.

2. Notation and definitions

This section contains notation and definitions that will be used throughout the rest of the paper.

2.1. Notation

Random variables in a Bayesian network will be denoted by capital letters, e.g., A, B, C. Sets of variables will be denoted by boldface capital letters, e.g., $\mathbf{X}$. All variables in this paper are assumed to take values in uncountable (continuous) state spaces. If $\mathbf{X}$ is a set of variables, $\mathbf{x}$ is a configuration of specific states of those variables. The continuous state space of $\mathbf{X}$ is denoted by $\Omega_{\mathbf{X}}$.

MTE probability potentials are denoted by lower-case greek letters, e.g., α, β, γ. In graphical representations, continuous nodes in Bayesian networks are represented by double-border ovals. Variables that are deterministic functions of their parents are represented by triple-border ovals. Shaded nodes are degenerate, indicating that evidence has restricted the variable to one value.

2.2. Conditional mass function (CMF)

When relationships between continuous variables are deterministic, the joint PDF does not exist. If Y is a deterministic relationship of variables in $\mathbf{X}$, i.e. $Y = g(\mathbf{X})$, the conditional mass function (CMF) for $\{Y|\mathbf{x}\}$ is defined as

$$p_{Y|\mathbf{x}} = \mathbf{1}\{y = g(\mathbf{x})\}, \tag{1}$$

where $\mathbf{1}\{A\}$ is the indicator function of the event A, i.e. $\mathbf{1}\{A\}(B) = 1$ if $B = A$ and 0 otherwise. Graphically, the conditionally deterministic relationship of Y given $\mathbf{X}$ is represented in a Bayesian network model as shown in Fig. 1, where $\mathbf{X}$ consists of a single continuous variable X.

2.3. Mixtures of truncated exponentials

A mixture of truncated exponentials (MTE) potential [9,12] has the following definition.

MTE potential. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be an n-dimensional random variable. A function $\phi : \Omega_{\mathbf{X}} \mapsto \mathscr{R}^+$ is an MTE potential if one of the next two conditions holds:

(1) The potential ϕ can be written as

$$\phi(\mathbf{x}) = a_0 + \sum_{i=1}^{m} a_i \exp\left\{\sum_{j=1}^{n} b_i^{(j)} x_j\right\} \tag{2}$$

for all $\mathbf{x} \in \Omega_{\mathbf{X}}$, where a_i, $i = 0, \ldots, m$ and $b_i^{(j)}$, $i = 1, \ldots, m, j = 1, \ldots, n$ are real numbers.

(2) The domain of the variables, $\Omega_{\mathbf{X}}$, is partitioned into hypercubes $\{\Omega_{\mathbf{X}}^1, \ldots, \Omega_{\mathbf{X}}^k\}$ such that ϕ is defined as

$$\phi(\mathbf{x}) = \phi_i(\mathbf{x}) \quad \text{if } \mathbf{x} \in \Omega_{\mathbf{X}_i}, \quad i = 1, \ldots, k, \tag{3}$$

where each ϕ_i, $i = 1, \ldots, k$ can be written in the form of Eq. (2).

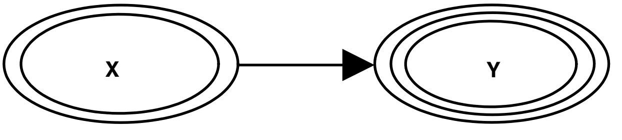

Fig. 1. Graphical representation of the conditionally deterministic relationship of Y given X determined by the CMF $p_{Y|x}$.

In the definition above, k is the number of *pieces* and m is the number of exponential *terms* in each piece of the MTE potential. In this paper, all MTE potentials are equal to zero in unspecified regions.

Moral et al. [10] proposes an iterative algorithm based on least-squares approximation to estimate MTE potentials from data. Moral et al. [11] describes a method to approximate conditional MTE potentials using a mixed tree structure. Cobb et al. [4] describes a nonlinear optimization procedure used to fit MTE parameters for approximations to standard PDFs, including the uniform, exponential, gamma, beta, and lognormal distributions.

Inference in continuous Bayesian networks where all conditional probability distributions are approximated by MTE potentials is performed using the operations of restriction, combination, and marginalization, as defined by Moral et al. [9] and further described by Cobb and Shenoy [1]. Restriction involves substituting real numbers representing evidence into a potential. Combination of two MTE potentials is pointwise multiplication. If two MTE potentials for $\mathbf{X}$ are denoted by ϕ and ψ, the combination of these two potentials is denoted by $\phi \otimes \psi$. Marginalization of a variable from an MTE potential is closed-form integration. If an MTE potential for $\mathbf{X}$ is denoted by ϕ, the marginalization of a variable $X \in \mathbf{X}$ from ϕ is denoted by ϕ^{-X}. Operations used to marginalize a variable from the combination of an MTE potential and a CMF are defined later in the paper.

3. Using CDFs to construct PDFs for deterministic variables

This section contains standard results from probability theory which describe methods of constructing CDFs and their corresponding PDFs for variables that are deterministic functions of their parents. These results are re-stated here for completeness.

3.1. Monotonically increasing functions

Consider a random variable Y which is a monotonically increasing deterministic function of a random variable X. A Bayesian network representing this relationship is shown in Fig. 1. The joint CDF for $\{X, Y\}$ represents the following probability:

$$\begin{aligned} F_{X,Y}(x,y) &= P[X \leqslant x, Y \leqslant y] \\ &= P[X \leqslant x]P[Y \leqslant y|X \leqslant x] \\ &= F_X(x)P[Y \leqslant y|X \leqslant x] \end{aligned}$$

for any $x \in \Omega_X$ such that $F_X(x) > 0$.

When Y is a monotonically increasing function of X, $X = g^{-1}(Y)$ and $P[Y \leqslant y|X \leqslant x] = 1$, thus $F_{X,Y}(x,y) = F_X(g^{-1}(y))$. Allowing x go to infinity in both sides of this expression gives $F_{X,Y}(\infty,y)$ or $F_Y(y) = F_X(g^{-1}(y))$. Differentiating both sides of this expression with respect to y (using the chain rule on the right-hand side) yields

$$f_Y(y) = f_X(g^{-1}(y))\frac{\mathrm{d}}{\mathrm{d}y}(g^{-1}(y)). \tag{4}$$

Thus, the Bayesian network where $X = g^{-1}(Y)$ and $f_Y(y)$ meets the above condition will have the same CDF of the original Bayesian network and the Bayesian networks are *equivalent*, as stated in the following proposition.

Proposition 1. *Suppose we have a Bayesian network with two variables X and Y with an arrow from X to Y, where Y is a conditionally deterministic, monotonically increasing function of X. Then, the equivalent Bayesian network with an arrow from Y to X, where X is a conditionally deterministic function of Y meets the conditions that* $f_Y(y) = f_X(g^{-1}(y))\frac{\mathrm{d}}{\mathrm{d}y}(g^{-1}(y))$ *and* $X = g^{-1}(Y)$.

When Y is a monotonically increasing (and therefore invertible) deterministic function of X, Proposition 1 gives a shortcut to finding the PDF of Y from the PDF of X that does not require the CDF of Y to be computed. We refer to using the operation in Proposition 1 as performing an "arc reversal" on the Bayesian network. After the operation is performed, the Bayesian network appears as in Fig. 2.

Example 1. Suppose that a random variable X has PDF

$$f_X(x) = \begin{cases} 3x^2 & \text{if } 0 < x < 1, \\ 0 & \text{elsewhere,} \end{cases}$$

and we want to find $f_Y(y)$ if $Y = g(X) = 4X^2$.

Note that $f_X(g^{-1}(y)) = 3y/4$ and $\frac{\mathrm{d}}{\mathrm{d}y}(g^{-1}(y)) = 1/(4\sqrt{y})$. Using Proposition 1, we compute

$$f_Y(y) = \begin{cases} \frac{3y}{4} \cdot \frac{1}{4\sqrt{y}} = \frac{3\sqrt{y}}{16} & \text{if } 0 < y < 4, \\ 0 & \text{elsewhere,} \end{cases}$$

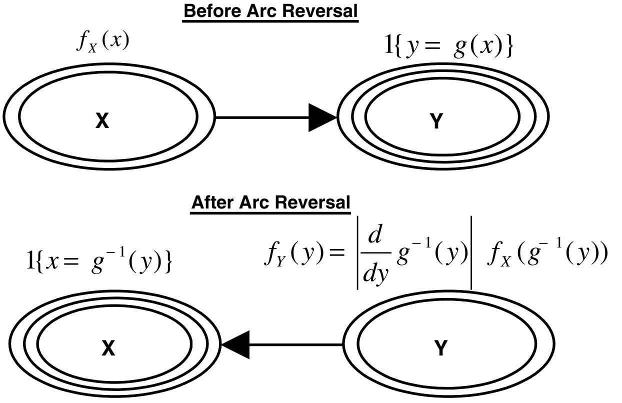

Fig. 2. Graphical representation of the conditionally deterministic relationship of X on Y before and after performing an "arc reversal" on the Bayesian network of Fig. 1.

3.2. Monotonically decreasing functions

Consider a random variable Y which is a monotonically decreasing function of a random variable X. A Bayesian network representing this relationship is shown in Fig. 1. The joint CDF for $\{X, Y\}$ represents the following probability:

$$\begin{aligned} F_{X,Y}(x,y) &= P[X \leqslant x, Y \leqslant y] \\ &= P[X \leqslant x] - P[x \leqslant g^{-1}(y)] \\ &= F_X(x) - F_X(g^{-1}(y)) \end{aligned}$$

for any $x \in \Omega_X$ such that $F_X(x) > 0$.

When Y is a monotonically decreasing function of X, $X = g^{-1}(Y)$, thus $F_{X,Y}(x,y) = F_X(x) - F_X(g^{-1}(y))$. Allowing x go to infinity in both sides of the last line of the expression above gives $F_{X,Y}(\infty, y)$ or $F_Y(y) = 1 - F_X(g^{-1}(y))$. Differentiating both sides of this expression with respect to y (using the chain rule on the right-hand side) yields

$$f_Y(y) = -f_X(g^{-1}(y))\frac{\mathrm{d}}{\mathrm{d}y}(g^{-1}(y)). \tag{5}$$

Thus, the Bayesian network where $X = g^{-1}(Y)$ and $f_Y(y)$ meets the above condition will have the same CDF of the original Bayesian network and the Bayesian networks are *equivalent*, as stated in the following proposition.

Proposition 2. *Suppose we have a Bayesian network with two variables X and Y with an arrow from X to Y, where Y is a conditionally deterministic, monotonically decreasing function of X. Then, the equivalent Bayesian network with an arrow from Y to X, where X is a conditionally deterministic function of Y meets the conditions that $f_Y(y) = -f_X(g^{-1}(y))\frac{\mathrm{d}}{\mathrm{d}y}(g^{-1}(y))$ and $X = g^{-1}(Y)$.*

When Y is a monotonically decreasing (and therefore invertible) deterministic function of X, Proposition 2 gives a shortcut to finding the PDF of Y from the PDF of X that does not require the CDF of Y to be computed. As in the monotonically increasing case, we refer to use of the operation in Proposition 2 as an arc reversal.

Example 2. Let X have the uniform PDF over the unit interval, i.e. $X \sim U(0,1)$. Find $f_Y(y)$ if $Y = g(X) = \frac{-\ln X}{\lambda}$.

Note that $f_X(g^{-1}(y)) = 1$ and $\frac{\mathrm{d}}{\mathrm{d}y}(g^{-1}(y)) = -\lambda \mathrm{e}^{-\lambda y}$. Using Proposition 2, we compute

$$f_Y(y) = \begin{cases} (-1)\cdot -\lambda \mathrm{e}^{-\lambda y} = \lambda \mathrm{e}^{-\lambda y} & \text{if } 0 < y < \infty, \\ 0 & \text{elsewhere}, \end{cases}$$

3.3. Linear CDF marginalization operator

Suppose Y is a conditionally deterministic linear function of X, i.e. $Y = g(X) = aX + b$, $a \neq 0$. The following operation will be used to determine the marginal PDF for Y:

$$f_Y(y) = (f_X \otimes p_{Y|x})^{-X}(y) = \frac{1}{|a|}\cdot f_X\left(\frac{y-b}{a}\right). \tag{6}$$

In this operation, X is marginalized from the combination of the PDF for X and the CMF for Y given X. The definition of the Linear CDF Marginalization Operator follows directly from the expressions in Propositions 1 and 2.

Example 3. Suppose that a random variable X has PDF

$$f_X(x) = \begin{cases} 6x(1-x) & \text{if } 0 < x < 1, \\ 0 & \text{elsewhere.} \end{cases}$$

Find $f_Y(y)$ if the deterministic relationship $Y = g(X) = 2X + 1$ is represented by the CMF $p_{Y|x} = \mathbf{1}\{y = 2x + 1\}$.

Note that $f_X(\frac{y-b}{a}) = f_X(\frac{y-1}{2}) = 6(\frac{y-1}{2}) \cdot (1 - (\frac{y-1}{2})) = -\frac{3}{2}y^2 + 6y - \frac{9}{2}$.

Using the operation in (6), we find the PDF for Y as

$$f_Y(y) = (f_X \otimes p_{Y|x})^{-X}(y) = \begin{cases} -\frac{3}{4}y^2 + 3y - \frac{9}{4} & \text{if } 1 < y < 3, \\ 0 & \text{elsewhere.} \end{cases}$$

The following theorem is required for inference using MTE potentials in Bayesian networks with linear conditionally deterministic variables.

Theorem 3. *If $\phi_1(x)$ is an MTE potential for X and Y is a conditionally deterministic linear function of X represented by the CMF $p_{Y|x}$, then $\phi_2(y) = (\phi_1 \otimes p_{Y|x})^{-X}(y)$ is an MTE potential.*

Proof. Multiplication by $1/a$ or $-1/a$ is multiplication by a constant. MTE potentials are closed under multiplication by constants (the constants a_0 and a_i, $i = 1, \ldots, m$ in (2) are revised). Exponential terms of the form $\exp\{x\}$ in $\phi_1(x)$ are revised to be of the form $\exp\{\frac{1}{a}y - \frac{b}{a}\}$ in $\phi_2(y)$. Since $\exp\{\frac{1}{a}y - \frac{b}{a}\} = \exp\{\frac{1}{a}y\} \cdot \exp\{-\frac{b}{a}\}$ and $\exp\{-\frac{b}{a}\}$ is a constant, the result is an MTE potential of the form in (2). $\square$

3.4. Method of convolutions

Let us consider a case where a conditionally deterministic variable has more than one parent. Let Z be a deterministic function of random variables X and Y, where X has PDF $f_X(x)$ and $\{Y|x\}$ has density $f_{Y|x}(y)$. A Bayesian network representation of this case is shown in Fig. 3.

Suppose $Z = g(X, Y)$ is invertible in Y. Then by arguments similar to those used in Propositions 1 and 2, we can show that the Bayesian network in Fig. 3 is equivalent to the Bayesian network in Fig. 4, where X has PDF $f_X(x)$, $\{Z|x\}$ has density

$$f_{Z|x}(z) = \left|\frac{\partial}{\partial z} g^{-1}(x,z)\right| \cdot f_{Y|x}(g^{-1}(x,z)),$$

and $Y = g^{-1}(X, Z)$ is a conditionally deterministic function of X and Z.

We can consider the Bayesian network in Fig. 4 as being the network obtained from the Bayesian network in Fig. 3 by reversing the arc (Y, Z). We can now compute the marginal density of Z as follows:

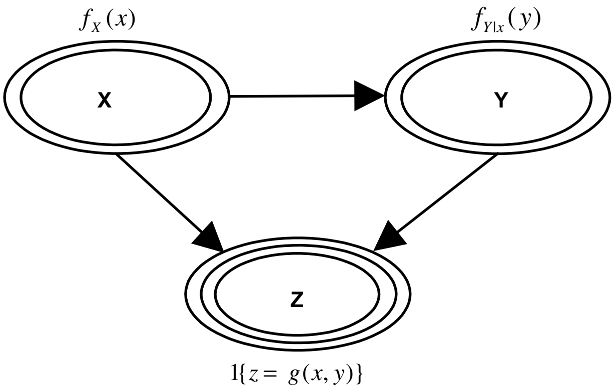

Fig. 3. Graphical representation of a Bayesian network where Z is a deterministic function of X and Y.

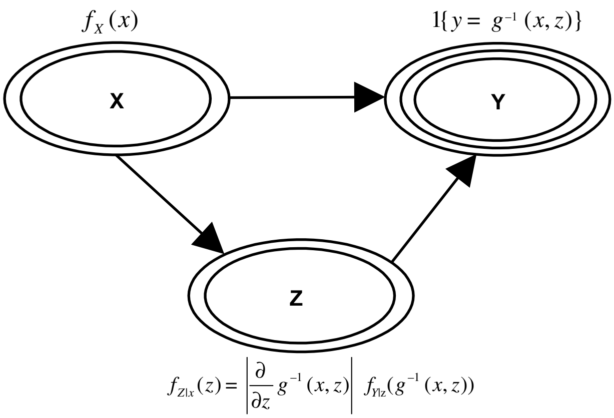

Fig. 4. Graphical representation of the Bayesian network in Fig. 3 after reversal of the arc (Y, Z).

$$
\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) \cdot f_{Z|x}(z)\, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \left(f_X(x) \cdot \left| \frac{\partial}{\partial z} g^{-1}(x,z) \right| \cdot f_{Y|x}(g^{-1}(x,z)) \right) \mathrm{d}x.
\end{aligned} \tag{7}
$$

The formula in (7) is called the *method of convolutions* in probability theory. The following theorem will be required for join tree operations when a variable is a linear conditionally deterministic function of its parents.

Proposition 4. *Let X and Y be continuous, possibly dependent random variables with joint PDF $f_{X,Y}$ and let $Z = a_1 \cdot X + a_2 \cdot Y + b$, $a_2 \neq 0$. The un-normalized joint PDF for $\{X, Z\}$ can be found as*

$$
f_{X,Z}(x,z) \propto f_{X,Y}\left(x, \frac{z - a_1 \cdot x - b}{a_2}\right).
$$

Proof. Follows directly from the method of convolutions in probability theory. □

We can replace X and a_1 in Proposition 4 with a vector of variables and a vector of non-zero constants, respectively, and the result holds. A transformation of the form in Proposition 4 is referred to as a *convolution* of the function $f_{X,Y}(x,y)$ [6]. To use the convolution formula in Proposition 4 to find PDFs for linear conditionally deterministic variables in hybrid Bayesian networks with MTE potentials, the following theorem is required.

Theorem 5. *If ϕ_1 is a joint MTE potential for $\{X, Y\}$ and $Z = a_1 \cdot X + a_2 \cdot Y + b$, $a_2 \neq 0$, the un-normalized joint PDF ϕ_2 for $\{X, Z\}$ calculated from the convolution of ϕ_1 is an MTE potential.*

Proof. Follows directly from the proof of Theorem 3. □

4. Join tree operations with linearly deterministic variables

Suppose we have a node in a join tree for a continuous Bayesian network containing a set of variables $\mathbf{X} = (X_1, \ldots, X_N)$. Assume a variable $X_i \in \mathbf{X}$ is a linear deterministic function of the remaining variables $\mathbf{X}' = \mathbf{X} \backslash X_i$, i.e.

$$X_i = g(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_N) = W + b,$$

where

$$W = a_1 \cdot X_1 + \cdots + a_{i-1} \cdot X_{i-1} + a_{i+1} \cdot X_{i+1} + \cdots + a_N \cdot X_N$$

with $a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_N$ and b defined as real numbers, with at least one of the slope coefficients in the linear equation not equal to zero. The joint PDF of $\mathbf{X}'$ is denoted by ϕ. The joint PDF for $\mathbf{X}$ does not exist; however, we can find the marginal PDF for X_i by using the operations defined in Section 3.

Consider the join tree in Fig. 5. The message passed from $\{\mathbf{X}\}$ to $\{\mathbf{X} \backslash X_k\}$ (where $X_k \in \mathbf{X}$ and $X_k \neq X_i$) is calculated using Proposition 4 as

$$\psi(\mathbf{x}'', x_i) = \left(\phi \otimes p_{X_i \backslash \mathbf{X}'}\right)^{-X_k}(\mathbf{x}'', x_i) = \phi\left(\left(x_i - \sum_{\substack{j=1 \\ j \notin \{i,k\}}}^{N} a_j x_j\right) \Big/ a_j\right),$$

where $\mathbf{x} = (\mathbf{x}'', x_i, x_k)$. The message from $\{\mathbf{X}\}$ to $\{\mathbf{X} \backslash X_k\}$ to $\{X_i\}$ is calculated as

$$\varphi(x_i) = \int_{\Omega_{\mathbf{X}''}} \psi(\mathbf{x}'', x_i)\, \mathrm{d}\mathbf{x}''.$$

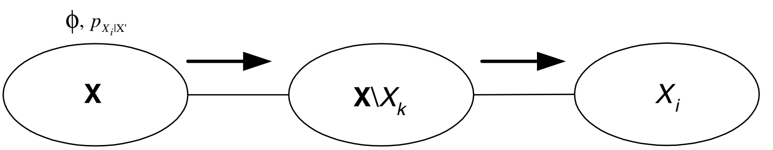

Fig. 5. A join tree for a Bayesian network with a deterministic variable.

The potential φ is normalized to find the posterior marginal for X_i. If ϕ is initially an MTE potential, φ is an MTE potential because the first message results in an MTE potential according to Theorem 5 and the second message results in an MTE potential because the class of MTE potentials is closed under marginalization.

The next example utilizes the MTE approximation to the normal PDF for join tree operations with a deterministic variable in order to compare answers to Hugin software.

Example 5. Consider the Bayesian network depicted in Fig. 6. Suppose $X \sim N(0,1)$, $Y \sim N(1,1)$, and Z is a conditionally deterministic function of its parents, $Z|x,y \sim N(2+x-y,0)$. We can calculate the marginal distribution of Z by passing messages in the join tree shown in Fig. 7.

The PDFs for X and Y, denoted by f_X and f_Y, respectively, are combined to form the joint PDF for $\{X,Y\}$ and sent to $\{X,Y,Z\}$ in the join tree. We next calculate the PDF for $Z = X - Y + 2$ using Proposition 4 as follows:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, x - z + 2)\,\mathrm{d}x.$$

Note that in this case, since X and Y are independent, $f_{X,Y}(x,y) = f_X(x) f_Y(y)$. Thus, if f_X and f_Y are maintained as decomposed potentials in the message from $\{X,Y\}$ to $\{X,Y,Z\}$ the calculation above can be simplified to

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(x - z + 2)\,\mathrm{d}x.$$

The marginal PDF for Z (shown in Fig. 8) was created by approximating the normal PDFs in this example with the MTE approximation to the normal PDF presented in Cobb and Shenoy [1]. The expected value and variance of this marginal PDF are 1.0000 and 1.9638. These answers are comparable with exact results obtained using Hugin software, which gives an expected value and variance of 1.0000 and 2.0000, respectively.

Suppose we obtain evidence that $Z = 3$ and pass this evidence as a message from $\{Z\}$ to $\{X,Y,Z\}$. Since the existing potential for Z states that $Z = 2 + X - Y$, the evidence dictates the new deterministic relationship $X = Y + 1$, which is expressed as the CMF $p_{X|y}$ and sent from $\{X,Y,Z\}$ to $\{X,Y\}$ in the join tree.

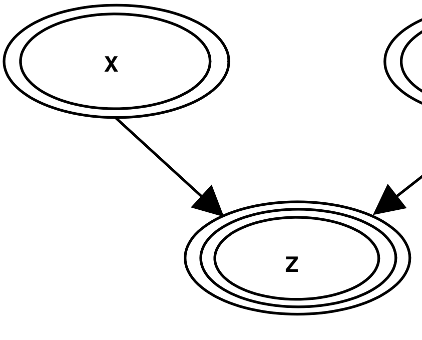

Fig. 6. The Bayesian network for Example 5.

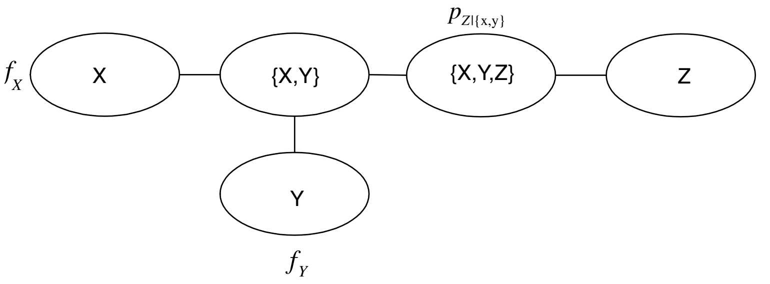

Fig. 7. The join tree for Example 5.

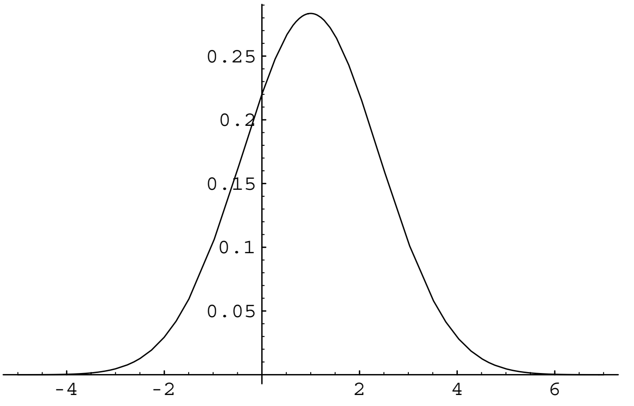

Fig. 8. The marginal PDF for Z in Example 5.

The variables X and Y are no longer independent and now have a linear conditionally deterministic relationship. The revised Bayesian network is depicted in Fig. 9. To calculate the revised marginal distribution for X we combine $f_X(x)$ with the distribution created by applying the linear CDF marginalization operator in (6) to the prior distribution for Y (the latter is the message from $\{X, Y\}$ to $\{X\}$) as follows:

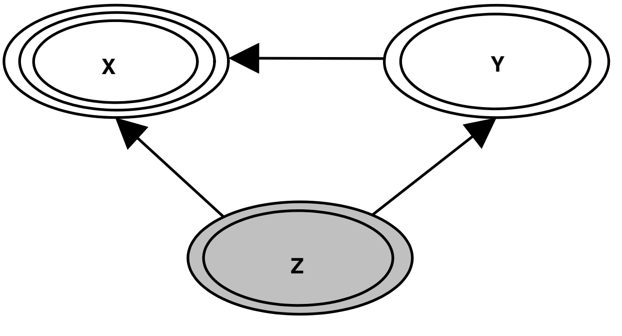

Fig. 9. The revised Bayesian network for Example 5 after observing evidence on Z.

$$f_{X_{ev}}(x) = K \cdot f_X(x) \cdot (f_Y \otimes p_{X|y})^{-Y}(x) = K \cdot f_X(x) \cdot f_Y(x-1).$$

In this calculation, K is a normalization constant. The expected value and variance of the posterior marginal PDF for X are calculated as 1.0000 and 0.5004, respectively. These answers are comparable with exact results obtained using Hugin software, which gives an expected value and variance of 1.0000 and 0.5000, respectively.

To calculate the revised marginal distribution for Y, we combine the prior distribution for Y with the distribution created by applying the linear CDF marginalization operator in (6) to the prior distribution for X (the latter is the message from $\{X, Y\}$ to $\{Y\}$) as follows:

$$f_{Y_{ev}}(y) = K \cdot f_Y(y) \cdot (f_X \otimes p_{Y|x})^{-X}(y) = K \cdot f_Y(y) \cdot f_X(y+1).$$

In this calculation, K is a normalization constant. Propositions 1 and 2 allow us to use either $p_{X|y}$ or $p_{Y|x}$ as equivalent expressions of the deterministic relationship between Y and X. The expected value and variance of the posterior marginal PDF for Y are calculated as 0.0000 and 0.5004, respectively. These answers are comparable with exact results obtained using Hugin software, which gives an expected value and variance of 0.0000 and 0.5000, respectively.

5. Example

The Bayesian network in this example (shown in Fig. 10) contains one variable (A) which follows a beta distribution, one variable (C) with a Gaussian potential, and one variable (B) which is a linear conditionally deterministic function of its parent. All probability potentials are approximated in the calculations by MTE potentials.

5.1. Representation

The probability distribution for A is a beta distribution with parameters $\alpha = 2.7$ and $\beta = 1.3$, i.e. £$(A) \sim Beta(2.7, 1.3)$. The PDF for A is approximated (using the methods described in [4]) by an MTE potential as follows:

$$\alpha(a) = P(A) = \begin{cases} -5.951669 + 5.573316\exp\{0.461388a\} & \\ \quad -0.378353\exp\{-6.459391a\} & \text{if } 0 < a < d^-, \\ 0.473654 - 6.358483\exp\{-2.639474a\} & \\ \quad +2.729395\exp\{-0.331472a\} & \text{if } d^- \leqslant a < m, \\ 1.823067 - (5.26\text{E}-12)\exp\{26.000041a\} & \\ \quad +0.035775\exp\{0.529991a\} & \text{if } m \leqslant a < 1, \\ 0 & \text{elsewhere.} \end{cases}$$

where $m = (1-\alpha)/(2-\alpha-\beta) = 0.85$ and

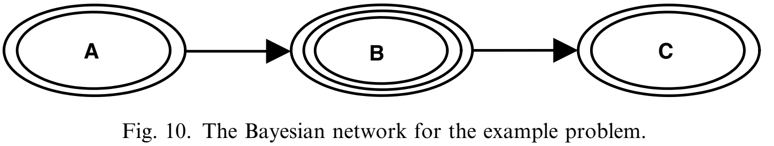

Fig. 10. The Bayesian network for the example problem.

$$d^{-} = \frac{(\alpha-1)(\alpha+\beta-3) - \sqrt{(\beta-1)(\alpha-1)(\alpha+\beta-3)}}{(\alpha+\beta-3)(\alpha+\beta-2)} = 0.493.$$

The MTE potential for A is shown graphically in Fig. 11, overlayed on the actual *Beta*(2.7, 1.3) distribution.

The probability distribution for B is defined as £$(B|a) \sim N(2a+1,0)$. The conditional distribution for B is represented by a CMF as follows:

$$\beta(a,b) = p_{B|a}(a,b) = \mathbf{1}\{b = 2a+1\}(a,b).$$

The probability distribution for C is defined as £$(C|b) \sim N(2b+1,1)$. This distribution is modeled with the MTE approximation to the normal PDF (denoted by δ) from [1].

5.2. Computing messages

The join tree for the example problem is shown in Fig. 12.

The messages required to calculate prior marginals for each variable in the network without evidence are as follows:

(1) α from $\{A\}$ to $\{A,B\}$
(2) $(\alpha \otimes \beta)^{-A}$ from $\{A,B\}$ to $\{B\}$ and $\{B\}$ to $\{B,C\}$
(3) $((\alpha \otimes \beta)^{-A} \otimes \delta)^{-B}$ from $\{B,C\}$ to $\{C\}$

5.3. Prior marginals

The prior marginal distribution for B is the message sent from $\{A,B\}$ to $\{B,C\}$. The expected value and variance of this distribution are calculated as 2.3488 and 0.1758, respectively. The prior marginal distribution for C is the message sent from $\{B,C\}$ to $\{C\}$. The expected value and variance of this distribution are calculated as 5.6975 and 1.6851, respectively. The prior marginal distributions for B and C are shown graphically in Fig. 13.

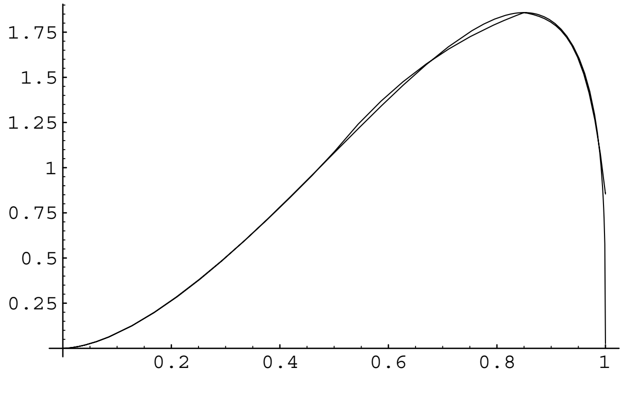

Fig. 11. The MTE potential for A overlayed on the actual *Beta*(2.7, 1.3) distribution.

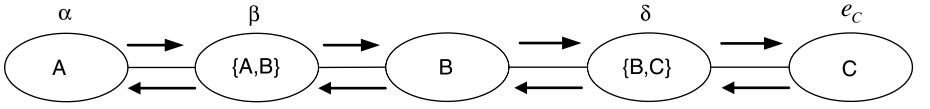

Fig. 12. The join tree for the example problem.

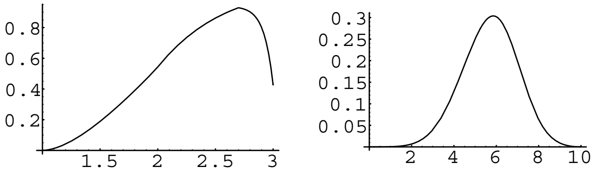

Fig. 13. The prior marginal distributions for B (left) and C (right).

5.4. Entering evidence

Assume evidence exists that $C = 6$ and define $e_C = 6$. Define $\eta = (\alpha \otimes \beta)^{-A}$ and $\vartheta(a, b) = p_{A|b}(a, b) = \mathbf{1}\{a = 0.5b - 0.5\}(a, b)$ as the potentials resulting from the reversal of the arc between A and B. The evidence $e_C = 6$ is passed from $\{C\}$ to $\{B, C\}$ in the join tree, where the existing potential is restricted to $\delta(b, 6)$. This likelihood potential is passed from $\{B, C\}$ to $\{B\}$ in the join tree.

Denote the un-normalized posterior marginal distribution for B as $\xi'(b) = \eta(b) \cdot \delta(b, 6)$. The normalization constant is calculated as $K = \int_b(\eta(b) \cdot \delta(b,6))\,\mathrm{d}b = 0.2344$. Thus, the normalized marginal distribution for B is found as $\xi(b) = K^{-1} \cdot \xi'(b)$. The expected value and variance of this distribution (which is displayed in Fig. 14) are calculated as 2.5049 and 0.0771, respectively.

Using the results of Proposition 1, we determine the posterior marginal distribution for A. Define $\theta = (\xi \otimes \nu)^{-B}$ as:

$$\theta(a) = \frac{1}{0.5}\xi(2a + 1).$$

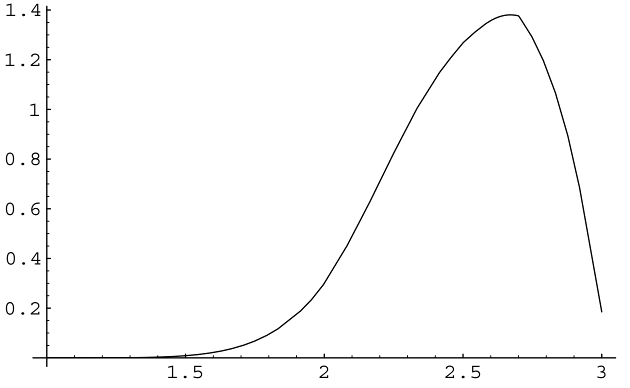

Fig. 14. The posterior marginal distribution for B considering the evidence $C = 6$.

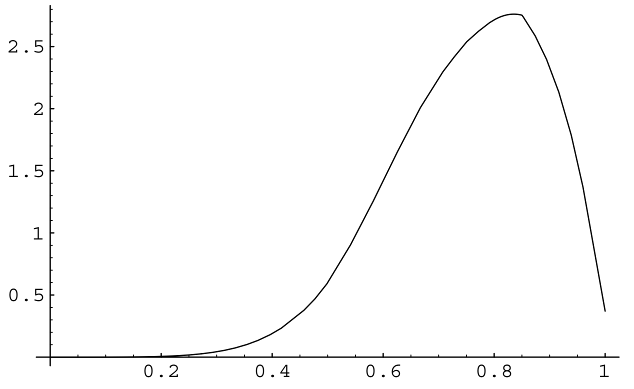

Fig. 15. The posterior marginal distribution for A considering the evidence ($c = 6$).

The CMF $v(a, b) = p_{B|a}(a, b) = \mathbf{1}\{b = 2a + 1\}(a, b)$ is obtained by reversing the arc between A and B in Fig. 10. The expected value and variance of this distribution are calculated as 0.7525 and 0.0193, respectively. The posterior marginal distribution for A considering the evidence is shown graphically in Fig. 15.

6. Summary and conclusions

This paper has described operations required for inference in continuous Bayesian networks containing variables that are linear conditionally deterministic functions of their parents. Since the joint PDF for a network with deterministic variables does not exist, the operations presented are derived from the method of convolutions in probability theory. Similar operations to those presented in this paper are incorporated in an inference algorithm for hybrid Bayesian networks (containing discrete and continuous variables) in [3]. This algorithm requires a "mixed potential" representation to accommodate mixed distributions. Nonlinear deterministic relationships can be accommodated in continuous Bayesian networks by extending the operations in this paper to piecewise linear functions which approximate nonlinear functions, as shown in [2].

Acknowledgments

This research was completed while the first author was a doctoral student at the University of Kansas and was supported by a research assistantship from the School of Business. We thank Thomas D. Nielsen, Antonio Salmerón, participants of the PGM 2004 conference, and two anonymous referees for valuable comments, suggestions, and discussions.

References

[1] B.R. Cobb, P.P. Shenoy, Inference in hybrid Bayesian networks with mixtures of truncated exponentials, Int. J. Approx. Reason., in press, doi:10.1016/j.ijar.2005.06.002.

[2] B.R. Cobb, P.P. Shenoy, Nonlinear deterministic relationships in Bayesian networks, in: L. Godo (Ed.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Lecture Notes in Artificial Intelligence, vol. 3571, Springer, Heidelberg, 2005, pp. 27–38.

[3] B.R. Cobb, P.P. Shenoy, Hybrid Bayesian networks with linear deterministic variables, in: F. Bacchus T. Jaakkola (Eds.), Uncertainty in Artificial Intelligence, vol. 21, AUAI Press, Corvallis, OR, 2005, pp. 136–144.
[4] B.R. Cobb, R. Rumí, P.P. Shenoy, Approximating probability density functions with mixtures of truncated exponentials, in: Proceedings of the Tenth Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU-04), 2004, pp. 429–436. Available from: <http://lark.cc.ku.edu/~pshenoy/Papers/IPMU04.pdf>.
[5] R.G. Cowell, A.P. Dawid, S.L. Lauritzen, D.J. Spiegelhalter, Probabilistic Networks and Expert Systems, Springer, New York, 1999.
[6] R.J. Larsen, M.L. Marx, An Introduction to Mathematical Statistics and its Applications, Prentice-Hall, Upper Saddle River, NJ, 2001.
[7] S.L. Lauritzen, Propagation of probabilities, means and variances in mixed graphical association models, J. Am. Stat. Assoc. 87 (1992) 1098–1108.
[8] S.L. Lauritzen, F. Jensen, Stable local computation with conditional Gaussian distributions, Stat. Comput. 11 (2001) 191–203.
[9] S. Moral, R. Rumí, A. Salmerón, Mixtures of truncated exponentials in hybrid Bayesian networks, in: P. Besnard, S. Benferhart (Eds.), Symbolic and Quantitative Approaches to Reasoning under Uncertainty, Lecture Notes in Artificial Intelligence, vol. 2143, Springer, Heidelberg, 2001, pp. 156–167.
[10] S. Moral, R. Rumí, A. Salmerón, Estimating mixtures of truncated exponentials from data, in: J.A. Gamez, S. Salmerón (Eds.), Proceedings of the First European Workshop on Probabilistic Graphical Models (PGM–02), Cuenca, Spain, 2002, pp. 135–143.
[11] S. Moral, R. Rumí, A. Salmerón, Approximating conditional MTE distributions by means of mixed trees, in: T.D. Nielsen, N.L. Zhang (Eds.), Symbolic and Quantitative Approaches to Reasoning under Uncertainty, Lecture Notes in Artificial Intelligence, vol. 2711, Springer, Heidelberg, 2003, pp. 173–183.
[12] R. Rumí, Modelos de Redes Bayesianas con Variables Discretas y Continuas, Doctoral Thesis, Universidad de Almería, Departamento de Estadística y Matemática Aplicada, Almería, Spain, 2003.